昆虫学报 *Acta Entomologica Sinica*, October 2017, 60(10): 1168 - 1177 doi: 10.16380/j.kcxb.2017.10.007

Mining the secretome of *Delia antiqua* (Diptera: Anthomyiidae) based on the transcriptomic and gene expression data and analysis of its potential roles in diapause development

HAO You-Jin*, XU Yan-Ling, CHEN Bin

(Chongqing Key Laboratory of Vector Insects, Institute of Entomology and Molecular Biology,
Chongqing Normal University, Chongqing 401331, China)

Abstract: 【Aim】Insect diapause is a critical eco-physiological adaptation for seasonal survival, but little is known about the molecular mechanisms of summer and winter diapause of *Delia antiqua*. Here, our main goal is to identify candidate genes linked to pupal diapause of *D. antiqua*. 【Methods】A straightforward *in silico* approach was performed to predict the secreted diapause-related proteins based on the RNA-seq and digital gene expression datasets. Amino acid sequence similarity and basic physiochemical characteristics analysis were conducted. Gene expression levels of 12 selected genes from RNA-seq data were further validated in non-diapause pupae, and different developmental stages of summer and winter diapause pupae by qRT-PCR. 【Results】A total of 38 putative secreted proteins were differentially regulated between the summer and winter diapause pupae of *D. antiqua*. Gene Ontology analysis revealed differentially expressed genes to be associated with odorant binding, chitin and lipid metabolic processes, innate immunity and development, *etc*. Among them, regulation of lipid mobilization and chitin-related metabolism was found to be important during diapause development. qRT-PCR analyses showed that the expression profiles of the 12 selected genes were in good agreement with results from RNA-seq ($R = 0.862$, $P = 0.001$). 【Conclusion】This study establishes a robust pipeline for the discovery of diapause-related genes by secretome mining based on the high-throughput sequencing.

Key words: *Delia antiqua*; diapause; secreted protein; transcriptome; digital gene expression; RNA-seq

1 INTRODUCTION

The onion fly, *Delia antiqua*, is an important pest and widely distributes in Europe, Asian, and North America (Hao *et al*., 2016) and can exploit summer or winter pupal diapause to survive the severe seasons. Both summer and winter diapause occurs when the head evagination is completed. It makes *D. antiqua* an ideal insect model to decipher and compare molecular mechanisms regulating summer diapause (SD) and winter diapause (WD). It has been known that diapause is characterized by a down-regulation in the expression of many genes as well as the up-regulation of others (Denlinger, 2002). Recent molecular studies have uncovered a variety of diapause-specific genes from various insect species (Qi *et al*., 2015, 2016; Hao *et al*., 2016; Meyers *et al*., 2016; Zhao *et al*., 2017). Yet, the precise molecular and physiological processes involved in the induction, maintenance and termination of diapause remain largely unknown. Therefore, it will be interesting to see whether *D. antiqua* exhibits a similar array of genes during its diapause.

During insect development, some secreted proteins are essential for cell communications within the organism (Horabin *et al*., 2003) or modulating responses to environmental stresses (Durand *et al*., 2010). Others have enzymatic functions involved in food degradation (Robich and Denlinger, 2005; Macrae, 2010), and body remodeling (Jia *et al*., 2014) during the life stage transition. In this study, putative secreted proteins were predicted based on the first transcriptome of *D. antiqua* and gene expression was analyzed through seven digital gene expression (DGE) datasets. Interestingly, some gene homologues potentially associated with odorant, chitin and lipid metabolic processes, innate

---

基金项目: 重庆市科委项目(cstc2016jcyjA0152); "两江学者"计划专项经费(20136666); 国家自然科学基金项目(31372265); 国际原子能机构 IAEA 研究项目(18268/R2)

作者简介: 郝友进, 男, 1973 年 7 月生, 山东莒县人, 博士, 教授, 研究方向为昆虫发育分子生物学, E-mail: haoyoujin@hotmail.com

* 通讯作者 Corresponding author, E-mail: haoyoujin@hotmail.com

收稿日期 Received: 2017-06-26; 接受日期 Accepted: 2017-08-07

immunity, *etc.*, were firstly identified and differentially expressed at least one diapause developmental phase. These findings serve as an important basis for understanding the molecular mechanisms which underlies the pupal diapause development of *D. antiqua*.

2 MATERIALS AND METHODS

2.1 Insect rearing

Delia antiqua were reared on artificial diets at 50% -70% relative humidity under a photoperiod of 16L: 8D at 20 ±0.5℃, and these insects were kept in the laboratory for successive generations and diapause pupae were generated as previously described (Fu *et al.*, 2016). Pupae were sampled according to the date (in days) after the pupation (D0), washed with ddH$_2$O, and then stored at -80℃.

2.2 *In silicon* secretome prediction

Nucleotide sequences were retrieved from the transcriptome data of *D. antiqua* (SRR916227) (Zhang *et al.*, 2014). Approaches to predict and annotate excretory proteins were performed by the following three steps (Fig. 1).

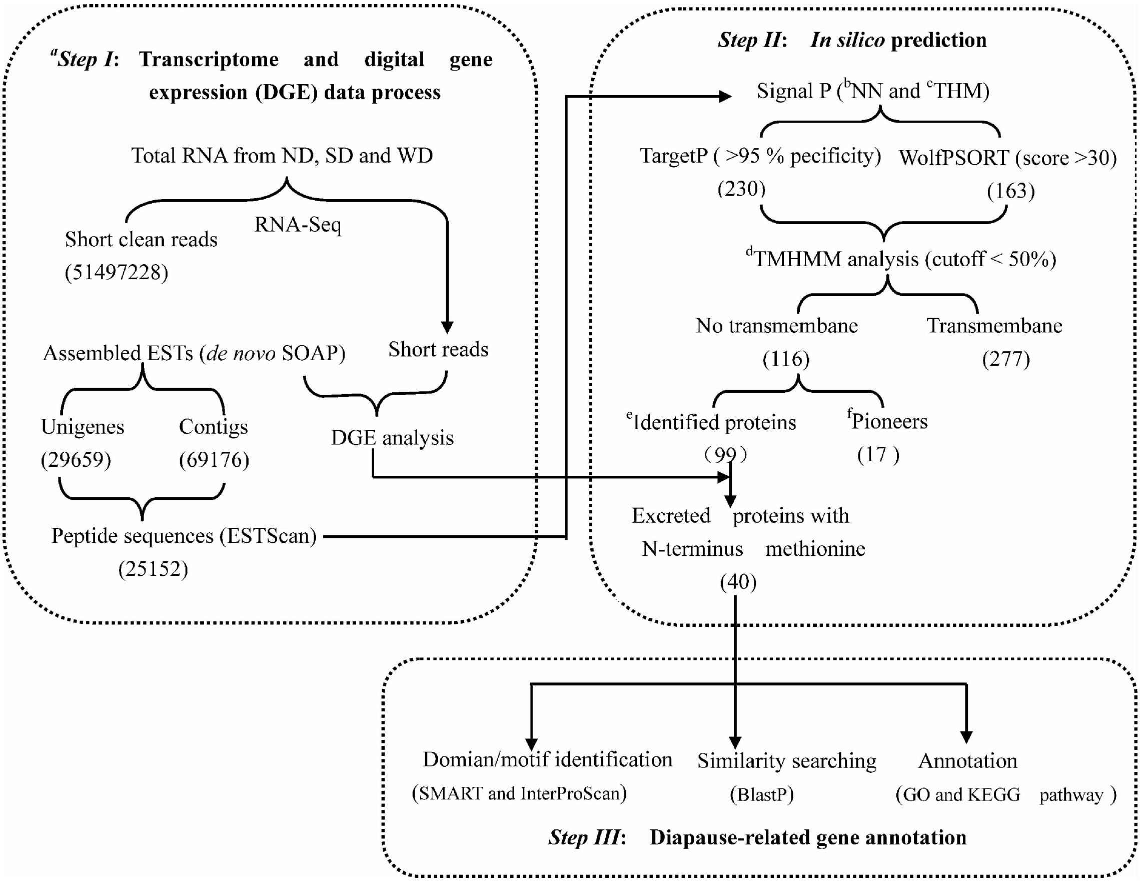
[a]Step I: Transcriptome and digital gene expression (DGE) data process
Total RNA from ND, SD and WD
RNA-Seq
Short clean reads (51497228)
Short reads
Assembled ESTs (de novo SOAP)
Unigenes (29659)
Contigs (69176)
DGE analysis
Peptide sequences (ESTScan) (25152)
Step II: In silico prediction
Signal P ([b]NN and [c]THM)
TargetP (>95 % pecificity) (230)
WolfPSORT (score >30) (163)
[d]TMHMM analysis (cutoff < 50%)
No transmembane (116)
Transmembane (277)
[e]Identified proteins (99)
[f]Pioneers (17)
Excreted proteins with N-terminus methionine (40)
Domian/motif identification (SMART and InterProScan)
Similarity searching (BlastP)
Annotation (GO and KEGG pathway)
Step III: Diapause-related gene annotation

Fig. 1 Bioinformatic flowchart for the secretome prediction of diapause-related genes in *Delia antique* based on transcriptome and digital gene expression datasets from non-diapause pupae (ND), summer diapause pupae (SD) and winter diapause pupae (WD)

The *in silico* analysis was performed with high stringent selection criteria. [a]Total number of short reads sequenced by next generation sequencing technique to explore onion maggot diapause mechanism. [b]NN = neural network. [c] HMM = hidden Markov models. [d]TMHMM = trans-membrane hidden Markov model. [e] Identified proteins are those proteins with N-terminus Met in Blast analysis. [f]Pineers are those proteins without N-terminus Met.

2.2.1 Transcriptome and digital gene expression (DGE) analysis: Transcriptome data and digital gene expression libraries were generated for non-diapause (ND), initial phase of summer and winter diapause (SI and WI), maintenance phase of summer and winter diapause (SM and WM), quiescence phase of summer and winter diapause (SQ and WQ). Sequencing and assembling, and annotation procedures were described in a previous study (Hao *et al.*, 2016). To create a list of differentially expressed genes with high confidence for further analysis, a stringent false discovery rate (FDR) (FDR≤0.001) cutoff and $|\log_2^{\text{RPKM Ratio}}| \geq 1$ were used as the threshold.

2.2.2 Excretory proteins prediction: Firstly, amino acid sequences retrieved from the transcriptomic data of *D. antiqua* were fed into Signal P (http://www.cbs.dtu.dk/services/SignalP/) to predict the presence and cleavage site of signal peptides. Sequences with N-terminal signal peptide were passed to Wolf PSORT (https://wolfpsort.hgc.jp/) and TargetP (http://www.cbs.dtu.dk/services/TargetP/) for subcellular location prediction. The obtained dataset was then analyzed by TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). To define a positive hit, the following criteria were used: (a) signal peptide predicted by neural network (NN) with the scores mean S and D; (b) signal peptide predicted by Hidden Markov models (HMM) considering the value of probability; (c) Wolf PSORT (score cut-off ≥30) and TargetP (score cut-off ≥95%); and (d) signal peptide cleavage site located 10 – 40 amino acids from the N-terminus. Furthermore, because TMHMM may not distinguish signal peptides from transmembrane domains, deduced proteins with single transmembrane domain with 70 amino acids of the N-terminus were also considered as potential secreted proteins.

2.2.3 Diapause-related gene identification and annotation: A unigene was considered to be a diapause-related gene when it produces a $P < 0.05$ using the *Chi*-square algorithm in DGE analysis. Amino acid sequence similarities with known secreted proteins were searched using BlastP. The candidate sequences without transmembrane segments were passed to further analysis. Protein domain and family classification were analyzed by SMART (http://smart.embl-heidelberg.de/) and Pfam (http://pfam.sanger.ac.uk/).

2.3 Digital gene expression validation by qRT-PCR

To validate gene expression patterns identified in RNA-seq analyses, qRT-PCR was performed for 12 selected genes in different function categories. Total RNA was extracted from one pupa/each sampling point with Trizol reagent (Invitrogen, USA) according to the manufacturer's instructions. One μg total RNA was reverse-transcribed to generate the first-strand cDNA using SuperScript Ⅱ reverse transcriptase (Invitrogen, USA). Each PCR reaction (20 μL in total) including 2 μL cDNA, 10 μL 2 × SYBR® Premix Ex Taq™ Ⅱ (Invitrogen, USA), 0.8 μL each of forward and reverse primer (10 μmol/L) was performed with the following conditions: 94℃ for 2 min, followed by 40 cycles of 95℃ for 15 s, 55 – 58℃ (dependent on gene of interest) for 15 s and 72℃ for 20 s. The specificity of the amplicons were verified by melting curve and the fold change was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Three biological replicates and three technical replicates were performed for each sample. The house-keeping gene *GADPH* was used as the internal control. The primers were designed based on the assembled nucleotide sequences in transcriptome data using the online server (http://www.genscript.com/) and listed in Table 1.

Table 1 The primers used for qRT-PCR

Genes	Forward primer (5′ – 3′)	Reverse primer (5′ – 3′)	Amplicon size (bp)
Destabilase	GCCTTTGTGAGGCAATTAGTGG	ACCGCCATCAGACCAAAAGG	101
Trypsin inhibitor	CGTTGGGCCCGAATTGTGAT	ATCATCTCTGGCAAACCCGT	113
Kazal type inhibitor	AAGGGCATGATTTGTGGTTCG	TTAGCAGGCTCCCTCCTTCT	123
Multiple domain protein	ATGAATACTTGCCCGGCTGT	GCAGCGTGGAAGTACAAACC	142
Major royal jelly protein	CTTCCGCGATGACAGCTCTT	GGATTCATGGGTCCTACGGC	138
Chitinase	TGAAATTGGGAGCACCGGCT	TGGACCTTGGAAGGCAGTACC	122
Hypothetical protein	ATTCTGCTGGTGCTTCACAC	TTGGCGTTATTGCCGTTTGG	109
Insect defensin A/B	GCATGGTCGCCATAAAAGAGC	ATTTCGGCACACGCAAACAG	139
Cathepsin L	TAATACGGGCAAGCCGTTCC	AACAGGAACCGCACACAGAT	125
Kunitz type inhibitor	GCAAAGGAGTATGTCGTGCC	ATTTTCGTTGCCATCACAGCC	101
Cathepsin D	TCCCATGGTAGGTGGCCAAT	CCATTTGGGCGATACGCAAG	131
Glycosyl hydrolase	AGCATTGCAACTTATGGCGT	TCATTCCAATGGCCAGGACC	130
TAG lipase	CGCCTTCTAGATACCGCCTG	AAATGCCAGCCGGATGAGTT	145

3 RESULTS

3.1 Assembly and extraction of transcriptome data

Initially, 51 497 228 short clean reads (4 634 750 520 bases) with GC content of 39.03% were generated by deep RNA-seq from seven samples (ND, SI, SM, SQ, WI, WM and WQ) of *D. antiqua*. After SOAP *de novo* assembly, a total of 69 176 contigs and 29 659 unigenes were obtained, conceptually translated into 25 152 ploy-peptide sequences by ESTScan and used for the subsequent analysis. Additionally, 7 digital gene expression datasets were produced and mapped on the reference transcriptome.

3.2 Secretome and diapause-related gene prediction

Firstly, 25 152 ploy-peptide sequences were analyzed by Signal P and generated 555 potentially secreted proteins. Then, 230 and 163 proteins were predicted as secreted proteins by TargetP and Wolf PSORT, respectively. A shared 150 sequences were passed to TMHMM for the prediction of transmembrane proteins. Thirty four proteins with one or more transmembrane fragments were removed from the secretory protein dataset. Ninety nine out of 116 proteins with N-terminal methionine were predicted as secreted proteins. Combined with digital gene expression data, finally 38 genes were predicted (Table 2).

3.3 Sequence characteristics

Five binding proteins including odorant binding proteins (OBPs), chemosensory protein (CSP) and chitin-binding proteins (Table 2) were identified, and sequence comparison showed that Ungiene 757 and Unigene 2347 had 73% and 65% identity to OBP of *Stomoxys calccitrans*, respectively. However, they shared very low identity between each other, implying they do not have a high conservation and may have different functions. Unigene 674 contains a conserved insect pheromone-binding domain (Pfam: PF03392) and has highly conserved identity above 65% to *Calliphora stygia* CSPs (AID61324).

Two insect defensin A/B homologues (Unigene103 and Unigene2131) were predicted and have 70% and 83% identity to their homologues of *Protophormia terraenovae* at the amino acid level, respectively. However, only 34.6% identity was shared between each other, suggesting that they do not display a high conservation. Domain analysis revealed that both have a typical insect defensin domain "knottins". Unigene780 encoding a lectin was significantly up-regulated in SM *vs* SI. Sequence analysis revealed that its full length amino acid sequence matched to C-type lectin of *Drosophila willistoni* (90% identity). Domain analysis showed that it contains a conserved carbohydrate- recognition domain (InterPro: IPR001304).

Alignment analysis revealed that Unigene201 shared 60% identity to a Kazal-type protease inhibitor of *Drosophila erecta* (EDV58866). While, Unigene828 appeared to be with full opening reading frame and shared 44% identity to a Kazal-type protease inhibitor of *Drosophila grimshawi* (EDV99366). Pairwise alignment revealed that they shared 37% identity to each other. In addition, Unigene1037 was predicted to be a kunitz-type inhibitor and shared 61% identity to the best-matched homologues (NP _ 608799) in *D. melanogaster* at the amino acid level. Unigene 832 encoding a protein containing a typical serpin domain (Pfam: Pf 00079) and the amino acid sequence had 67% identity to serpins of *D. melanogaster*. Interestingly, a cysteine protease inhibitor Unigene597 was identified and showed 54% identity to sarcocystatin of *Sarcophaga crassipalpis* (P31727). A conserved cystatin domain QxVxG (x can be of several amino acids) and a potential target protease binding sit G were found in amino acid sequence.

Cathepsin D had 89% identity to their best hits in *Musca domestica* (ABM69086), while cathepsin L shared 91% identity to its homologue in *Sarcophaga peregrina* (BAA76272). Sequence analysis revealed that both cathepsin D and cathepsin L consist of a signal peptide, a propeptide inhibitor domain (I29) and a mature protein sequence.

Proteins with putative functions differing from the above categories were grouped as proteins with energy homeostasis and miscellaneous functions. Transcripts without similarity to known genes were grouped as hypothetical proteins. A transcript (Unigene 2646) encoding a major royal jelly protein was identified and it had 75% identity to a homologue of *D. persimilis* (EDW37044). Similarity searching revealed that the full-length of Ungiene5734 encodes a triacylglycerol lipase (TAG lipase) with 64% identity to its homologue of *D. willistoni* (EDW76570). Protein domain analysis found an α/β-hydrolase lipase domain (Pfam: PF04083) and an α/β-hydrolase domain (Pfam: PF00561). Unigene1435 encodes a multiple-domain protein including peritrophin A domain, low density lipoprotein receptor domain (LDLR), and chitin deacetylase-like domain. Similarity searching revealed that it has 92% identity to its homologues in *D. mojavensis* (EDW18929).

3.4 Gene expression pattern analysis

To gain insights into the global transcriptional patterns during the pupal diapause process, DGEs were identified from the normalized data by pairwise comparison of initiation phase of summer diapause *vs* non-diapause (SI *vs* ND) or initiation phase of winter diapause *vs* non-diapause (WI *vs* ND), maintenance phase *vs* initiation phase (SI *vs* SM,

Table 2 Identification and gene expression patterns of the putative excreted proteins

Gene ID	Homologous	GenBank accession no.	Species	E-value	Sequence identity	Gene expression level (absolute value of log$_2$ $^{\text{RPKM ratio}}$)					
						SI *vs* ND	SI *vs* SM	SM *vs* SQ	WI *vs* ND	WI *vs* WM	WM *vs* WQ
	Binding proteins										
Unigene757	Odorant binding protein	ADG96056	*Stomoxys calcitrans*	3E-72	104/142 (73%)	-	1.6↑	-	-	3.0↑	-
Unigene2347	Odorant binding protein	ADG96058	*Stomoxys calcitrans*	1E-58	91/140 (65%)	1.4↓	1.7↑	-	2.6↑	3.9↑	1.8↓
Unigene674	Chemosensory protein	AID61324	*Calliphora stygia*	2E-62	95/129(74%)	-	1.8↑	-	2.5↓	-	1.8↑
CL1512. Contig1	Chitin-binding protein	EDW72934	*Drosophila willistoni*	2E-48	75/101 (74%)	3.2↓	-	-	2.7↓	4.0↑	1.9↑
Unigene939	Peritrophin A	AAL15463	*Lucilia cuprina*	9E-33	70/153 (46%)	-	-	-	1.2↓	1.5↑	1.2↓
	Immunity										
Unigene103	Insect defensin A/B	CAA39152	*Protophormia terraenovae*	3E-40	66/94 (70%)	-	-	-	1↓	-	1↑
Unigene2131	Insect defensin A/B	CAA39152	*Protophormia terraenovae*	1E-52	78/94 (83%)	1.5↓	-	-	2.8↓	-	3↑
Unigene780	C-type lectin	EDW75548	*Drosophila willistoni*	8E-153	212/236 (90%)	-	1.6↑	-	1.6↓	1↑	-
CL481. Contig2	Destabilase	EDW32304	*Drosophila persimilis*	4E-70	103/152 (68%)	1.1↑	2.5↑	1↑	5.5↓	3.1↑	-
Unigene5056	Immunoglobulin-like	GI16817	*Drosophila mojavensis*	2E-108	155/274(57%)	3.1↑	-	-	-	2.0↓	-
	Protease inhibitors										
Unigene201	Kazal type inhibitor	EDV58866	*Drosophila erecta*	5E-26	44/73 (60%)	-	-	-	-	-	1↑
Unigene828	Kazal type inhibitor	EDV99366	*Drosophila grimshawi*	1E-15	35/79 (44%)	1.3↓	-	1.1↓	-	1.1↑	-
Unigene1037	BPTI/Kunitz type inhibitor	NP_608799	*Drosophila melanogaster*	2E-32	50/82(61%)	-	-	1↑	-	-	-
Unigene2593	BPTI/Kunitz type inhibitor	EDW59902	*Drosophila virilis*	5E-41	68/112 (61%)	1.2↑	-	-	3.2↓	2.0↓	1↑
Unigene832	Serpin	AAM52725	*Drosophila melanogaster*	0	262/390 (67%)	2.6↑	1.2↑	-	2.3↑	2.9↓	-
Unigene1097	Trypsin inhibitor	EDW43779	*Drosophila sechellia*	8E-10	34/87 (39%)	-	-	1.2↑	1.9↓	3.22↑	-
Unigene597	Cystatin	P31727	*Sarcophaga peregrina*	4E-41	67/124(54%)	-	-	-	-	2.1↑	1.1↓
	Proteases										
Unigene3614	Chitinase	AEN03038	*Bactrocera dorsalis*	0	343/478 (72%)	-	1.5↑	-	-	1↑	-
Unigene941	Cathepsin D	ABM69086	*Musca domestica*	0	334/374 (89%)	-	1.9↑	-	-	2.7↑	-
Unigene1230	Cathepsin L	BAA76272	*Sarcophaga peregrina*	0	501/550 (91%)	-	-	-	1.4↓	-	-

Table 2 continued

Gene ID	Homologous	GenBank accession no.	Species	E-value	Sequence identity	Gene expression level (absolute value of $\log_2^{\text{RPKM ratio}}$)					
						SI *vs* ND	SI *vs* SM	SM *vs* SQ	WI *vs* ND	WI *vs* WM	WM *vs* WQ
Unigene4495	Glycosyl hydrolase	EAL25033	*Drosophila pseudoobscura pseudoobscura*	0	304/393 (77%)	–	–	–	–	2.6↑	–
CL297. Contig1	Aminopeptidase	AAM77681	*Lucilia cuprina*	0	611/914(67%)	–	–	2.6↑	1.9↓	2.5↑	–
Unigene633	Cyclophilin	GH20330	*Drosophila grimshawi*	2E – 127	178/203(88%)	1↓	–	–	1.6↑	2.1↑	1.3↓
	Energy homeostasis										
Unigene2646	Major royal jelly protein	EDW37044	*Drosophila persimilis*	0	326/432 (75%)	–	2.2↑	1↑	1.4↑	2.1↑	–
Unigene5734	TAG lipase	EDW76570	*Drosophila willistoni*	0	259/406 (64%)	–	2.5↑	–	–	2.7↑	1.1↑
	Miscellaneous proteins										
Unigene1099	Transferrin	BAA06067	*Sarcophaga peregrina*	1E – 130	195/216 (90%)	–	–	–	–	1.3↓	–
Unigene3674	Imaginal disk growth factor	Q2PQM6	*Glossina morsitans morsitans*	0	253/423(60%)	2.8↓	1.1↑	–	2.8↓	4.1↑	2↓
Unigene1435	Multiple domain protein	EDW18929	*Drosophila mojavensis*	0	497/538 (92%)	2.4↑	–	1.1↓	2.4↑	2.1↓	–
CL2193. Contig1	ChBD2 + LDLa	ABI95429	*Lucilia cuprina*	0	503/533 (94%)	1.8↑	–	–	–	–	–
Unigene711	Ferritin heavy chain	ADI87390	*Lucilia sericata*	5E – 86	124/142(87%)	–	–	–	–	1.4↓	–
Unigene4795	Tweedle E	NP_609157	*Drosophila melanogaster*	3E – 80	146/169(86%)	1.5↓	3.1↑	1.4↓	1.7↓	1.3↑	2.2↑
Unigene1630	Conserved secreted protein	ADD20601	*Glossina morsitans morsitans*	3.00E – 54	81/130 (62%)	2↑	1↓	–	1.8↑	2.91↓	1↓
Unigene3981		EDW18799	*Drosophila mojavensis*	2E – 38	111/265 (42%)	1.9↓	1.7↑	2.4↑	–	4.4↑	1.1↓
Unigene7142		GJ13481	*Drosophila virilis*	2E – 41	75/106(71%)	2.1↑	–	–	3.2↑	2.0↓	–
Unigene957		ADD20240	*Glossina morsitans morsitans*	7E – 77	117/151(77%)	1.2↓	1.3↑	–	–	2.4↑	–
Unigene2031		CG30101	*Drosophila melanogaster*	1E – 154	272/324(84%)	–	2.2↑	–	1.8↓	2.3↑	–
Unigene844		GF21087	*Drosophila ananassae*	2E – 18	58/174(33%)	2↑	2.5↑	1.6↑	7.7↑	1.2↑	3.2↑
Unigene1510		GK15109	*Drosophila willistoni*	1E – 19	45/118(38%)	1.9↓	–	–	–	2.3↑	–

Gene expression fold up-regulated (↑) or down-regulated (↓); A fold change less than 2 is marked with dash (–); ND: Non-diapause; SI: Initial phase of summer diapause; SM: Maintenance phase of summer diapause; SQ: Quiescence phase of summer diapause; WI: Initial phase of winter diapause; WM: Maintenance phase of winter diapause; WQ: Quiescence phase of winter diapause.

WM *vs* WI), and quiescence phase *vs* maintenance phase (SQ *vs* SM, WQ *vs* WM). The fold changes were shown in Table 2. qRT-PCR analysis revealed showed that the 12 selected genes were differentially expressed at least one diapause phase and showed a very similar expression trend as in RNA-seq analysis with a spearman correlation of $R = 0.862$ ($P = 0.001$) (Fig. 2).

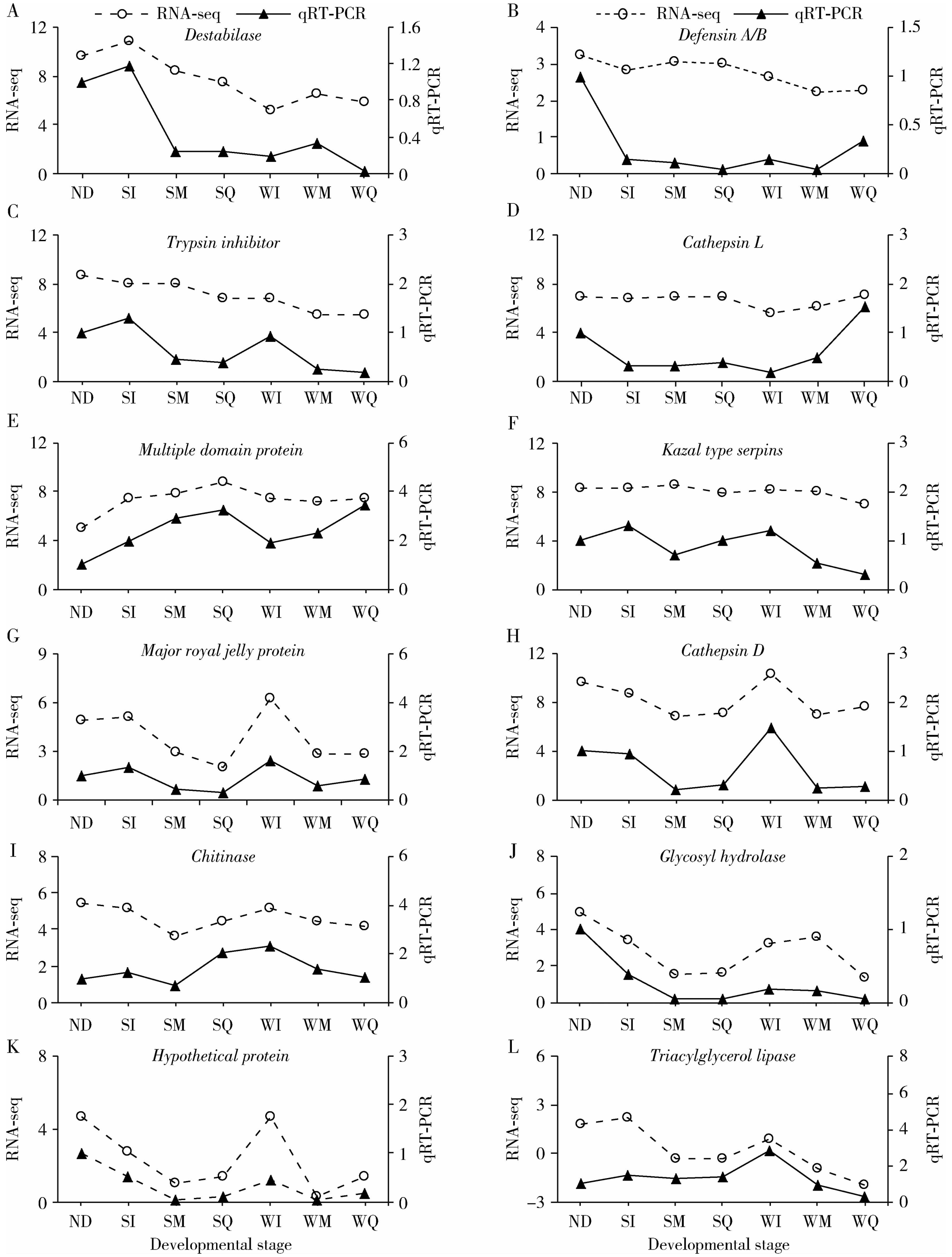

Fig. 2 Validation of RNA-seq results by qRT-PCR

Expression levels of 12 selected genes in seven samples (ND, SI, SM, SQ, WI, WM and WQ) were detected by qRT-PCR. ND: Non-diapause; SI: Initial phase of summer diapause; SM: Maintenance phase of summer diapause; SQ: Quiescence phase of summer diapause; WI: Initial phase of winter diapause; WM: Maintenance phase of winter diapause; WQ: Quiescence phase of winter diapause. Relative expression levels were derived from three biological replicates in qRT-PCR using ND as the calibrator ($2^{-\Delta\Delta Ct}$ method), and log$_2$ values of normalized RPKM of RNA-seq results.

The transcripts of 9 genes were significantly up-regulated in SI *vs* ND (Table 2). The highest changed expression level was 8.6-fold for *tweedle E*, then 5.7-fold for *destabilase*, *triacylglycerol lipase* and new hypothetical protein (Unigene 844). The transcripts of *aminopeptidase* (6.5-fold) and two hypothetical proteins (5.3-fold for Unigene 3981, and 3-fold for Unigene 844) were significantly higher at SM than at SQ. Whereas 8 genes were up-regulated in WI *vs* ND (Table 2). The expression level of a hypothetical gene (Ungiene 844) was increased 59.2-fold in WI *vs* ND, followed by Unigene 3981 (19.4-fold), OBP (Unigene 2347) (15.2-fold), *cathepsin D* and *cathepsin L* (7.3-fold), and major royal jelly protein (4.4-fold). In contrast, only 8 genes were up-regulated in maintenance phase compared with the quiescence phase. The most abundant transcripts were Unigene 844 (9.2-fold), *defensin A/B* (Unigene 2131, 8-fold) and *tweedle E* (4.6-fold).

4 DISCUSSION

Based on a straightforward *in silico* approach, we generated the first comprehensive overview of the secreted proteins from the diapause onion fly pupa. Thirty eight genes were identified and potential functions of key genes during the diapause development were discussed as below.

Insect OBPs are small, water-soluble proteins expressed in both olfactory and gustatory sensilla, as well as in other specialized tissues, and are secreted into the extracellular fluid. Generally, they were thought to be soluble hydrophilic odorants and bound by the respective receptors (Andronopoulou *et al.*, 2006, Mitaka *et al.*, 2011). Besides contributing to the recognition of odorants in insect, they also function as carriers in other developmental and physiological processes (Foret and Maleszka, 2006). As to the specific function of CSP and OBPs in the onion fly diapause pupae, we can only speculate. Perhaps the most intriguing hypothesis is that they might somehow regulate the chemosensory response to pheromonal odors in a hormonally and perhaps circadian sensitive manner. These findings served an important basis for understanding the molecular mechanisms of pheromones during the pupal diapause.

Previous study also showed that defensin was up-regulated during larval diapause in spruce budworm *Choristoneura fumiferana* (Palli *et al.*, 2001), but not in diapausing fly *D. triauraria* (Daibo *et al.*, 2001). In this study, the transcriptional level of *defensin A/B* (Unigene 2131) was 8-fold higher in WQ *vs* WM. In contrast, it was down-regulated 7-fold in WI *vs* ND, and 13-fold in WI *vs* WM. This distinct expression patterns suggest that it could help winter diapause pupa increase ability to withstand attaches by pathogens or paly a crosstalk function of diapause development. The *Drosophila* gene *fw* (C-lectin orthologue) mutation could severe impair eye and bristle development (Leshko-Lindsay and Corces, 1997). In flesh fly *S. peregrina*, a C-lectin was found to play roles in imaginal disc differentiation (Kawaguchi *et al.*, 1991). Another insect C-lectin (designated as regenectin) was expressed selectively during cockroach tissue regeneration and proposed to play a potential role in control of development (Arai *et al.*, 1998). The down-regulated expression pattern indicated that *D. antiqua* C-lectin might function directly or indirectly in immune defense and play a role in the development during the diapause. Destabilase is a member of lysozyme superfamily and its activity leads to the dissolution of stabilized fibrin. For the first time, a transcript coding destabilase was identified in diapause insect and significantly expressed at WI (Fig. 2). Due to its uncharacterized biochemical activities, it is still unknown whether this homologue has the destabilase activity or not and what is the real function in the winter diapause initiation.

Two cystatins named *Scys-A* and *-B* have been previously described in the flesh fly *S. crassipalpis*. *Scys-A* was down-regulated throughout pupal diapause (Goto *et al.*, 2002). However, *Scys-B* was highly expressed in early diapause. In this context, authors hypothesized that *Scys-B* would play a role in halting development. Interestingly, a similar expression pattern of *Scys-B* was observed for Unigene597. Its transcript was significantly increased at WI, suggesting that it would have potential role in initiating the developmental arrest. Being no cystatins have been biochemically characterized in diapause insects, what possible role it may play in *D. antiqua* is unclear. Serine proteases are involved in a variety of physiological functions such as development, digestion and stress responses. Loss of these critical protease inhibitors results in death of the organism or death. The up-regulation of a serpin (HiD5) of the horn fly, *Haematobia irritans* has been observed and was thought to play an important role in inhibiting the differentiation and destruction of imaginal tissues during diapause (Kalischuk-Tymensen, 2001). On the contrary, a serpin was down-regulated at the maintenance phase of the Colorado potato beetle, *Leptinotarsa decemlineata* (Yocum *et al.*, 2009). Here, *D. antiqua* pupae ceases of feeding, the diverse expression patterns of serpins between diapause and non-diapause suggests that they play some roles except the inhibiting the activity of digestive proteases. To our knowledge, serpins in diapause insects have not been functionally studied. What possible role serpins may play is unclear. Therefore, further experiments should be designed to investigate its real function in the pupal diapause of *D. antiqua*.

Chitinases are members of the *O*-glycoside hydrolase superfamily and have been detected in molting fluid and gut tissues. In this study, Unigene 3614 showed a U-shaped expression pattern in summer diapause pupae. However, its expression kept approximately constant across the whole winter diapause period (Fig. 2). As a state of developmental arrest, the lower expression of chitinase in diapause pupae of *D. antiqua* is understandable; however, its higher expression level at the quiescence phase suggests that it is associated with the termination of summer diapause and the preparation for adult eclosion.

Imaginal disc growth factors (IDGFs) are a group of V chitinase-like proteins and have been shown to regulate cell proliferation and remodeling in insect. In *D. melanogaster*, four insulin-dependent *IDGFs* were found to promote cell proliferations (Kawamura *et al.*, 1999). It was also identified in *H. armigera* and found to be associated with the pupal diapause (Chen *et al.*, 2010). The highest expression of *D. antiqua IDGF* in ND pupae suggests that it plays a similar role in cell proliferations of imaginal disc. The decreased expression in the whole summer diapause period, WI and WM, and the up-regulated expression at WQ proved the arrested development in diapause process and development recovery in the quiescence phase.

In this study, excreted proteins were predicted and their gene expression patterns were compared between different developmental stages to know about the putative transcripts involved in summer diapause and winter diapause. Many putative excreted protein genes showed no obvious similarity or convergence in two diapause types. More differentially expressed genes were found in maintenance phase of winter diapause than in summer diapause. Regulation of lipid mobilization and chitin-related metabolism at the transcriptional level was found to be important during diapause development and seasonal variation (summer or winter). The differences in the types and quantities of putative genes suggest that the evolution of two diapause types has led to the diversity of gene function adapting to specific seasonal environments. Taken together, our results provide a start point for the understanding of how excreted proteins function in the diapause development of onion fly.

References

Andronopoulou E, Labropoulou V, Douris V, Woods DF, Biessmann H, Iatrou K, 2006. Specific interactions among odorant-binding proteins of the African malaria vector anopheles gambiae. *Insect Mol. Biol.*, 15(6): 797 - 811.

Arai T, Kawasaki K, Kubo T, Natori S, 1998. Cloning of cDNA for regenectin, a humoral C-type lectin of *Periplaneta americana*, and expression of the regenectin gene during leg regeneration. *Insect Biochem. Molec. Biol.*, 28(12): 987 - 994.

Chen L, Ma W, Wang X, Niu C, Lei C, 2010. Analysis of pupal head proteome and its alteration in diapause pupae of *Helicoverpa armigera*. *J. Insect Physiol.*, 56(3): 247 - 252.

Daibo S, Kimura MT, Goto SG, 2001. Upregulation of genes belonging to the drosomycin family in diapausing adults of *Drosophila triauraria*. *Gene*, 278(1 - 2): 177 - 184.

Denlinger DL, 2002. Regulation of diapause. *Annu. Rev. Entomol.*, 47(1): 93 - 122.

Durand N, Carot-Sans G, Chertemps T, Bozzolan F, Party V, Renou M, Debernard S, Rosell G. Maibeche-Coisne, M, 2010. Characterization of an antennal carboxylesterase from the pest moth *Spodoptera littoralis* degrading a host plant odorant. *PLoS ONE*, 5(11): e15026.

Foret S, Maleszka R, 2006. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res.*, 16(11): 1404 - 1413.

Fu DY, Si FL, Chen B, Xu YL, Hao YJ, 2016. Transcriptome-guided identification of the cuticular protein gene family of *Delia antiqua* (Diptera: Anthomyiidae) and expression analysis under non- and winter-diapause conditions. *Acta Entomol. Sin.*, 59(2): 172 - 184. [付丹影, 司风玲, 陈斌, 徐燕玲, 郝友进, 2016. 基于转录组的葱蝇表皮蛋白基因家族的鉴定及在非滞育和冬滞育条件下的表达分析. 昆虫学报, 59(2): 172 - 184]

Goto SG, Denlinger DL, 2002. Genes encoding two cystatins in the flesh fly *Sarcophaga crassipalpis* and their distinct expression patterns in relation to pupal diapause. *Gene*, 292(1 - 2): 121 - 127.

Hao YJ, Zhang YJ, Si FL, Fu DY, He ZB, Chen B, 2016. Insight into the possible mechanism of the summer diapause of *Delia antiqua* (Diptera: Anthomyiidae) through digital gene expression analysis. *Insect Sci.*, 23(3): 438 - 451.

Horabin JI, Walthall S, Vied C, Moses M, 2003. A positive role for patched in hedgehog signaling revealed by the intracellular trafficking of sex-lethal, the *Drosophila* sex determination master switch. *Development*, 130: 6101 - 6109.

Jia QQ, Liu Y, Liu HH, Li S, 2014. Mmp1 and Mmp2 cooperatively induce *Drosophila* fat body cell dissociation with distinct roles. *Sci. Rep.*, 4: 7535.

Kalischuk-Tymensen L, 2001. Isolation and Identification of Genes Expressed during Diapause in Horn Fly, *Haematobia irritans* (L). PhD Dissertation, Faculty of Arts & Science, University of Lethbridge, Lethbridge, Alberta.

Kawaguchi N, Komano H, Natori S, 1991. Involvement of *Sarcophaga* lectin in the development of imaginal discs of *Sarcophaga peregrina* in an autocrine manner. *Dev. Biol.*, 144(1): 86 - 93.

Kawamura K, Shibata T, Saget O, Peel D, Bryant PJ, 1999. A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. *Development*, 126(2): 211 - 219.

Leshko-Lindsay LA, Corces VG, 1997. The role of selectins in *Drosophila* eye and bristle development. *Development*, 124(1): 169 - 180.

Livak KJ, Schmittgen TD, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, 25(4): 402 - 408.

Macrae TH, 2010. Gene expression, metabolic regulation and stress tolerance during diapause. *Cell. Mol. Life Sci.*, 67(14): 2405 - 2424.

Meyers PJ, Powell TH, Walden KK, Schieferecke AJ, Feder JL, Hahn DA, Robertson HM, Berlocher SH, Ragland GJ, 2016. Divergence of the diapause transcriptome in apple maggot flies: winter regulation and post-winter transcriptional repression. *J. Exp. Biol.*, 219(17): 2613 - 2622.

Mitaka H, Matsuo T, Miura N, Ishikawa Y, 2011. Identification of odorant-binding protein genes from antennal expressed sequence tags of the onion fly, *Delia antiqua*. *Mol. Biol. Rep.*, 38(3): 1787 - 1792.

Palli SR, Kothapalli R, Feng Q, Ladd T, Perera SC, Zheng SC, Gojtan K, Pang ASD, Primavera MJ, Tomkins WL, Retnakaran A, 2001. Molecular analysis of overwintering diapause. In: Denlinger DL, Giebultowicz JM, Saunders DS eds. Insect Timing: Circadian Rhythmicity to Seasonality. Elsevier Science Inc., Amsterdam, The Netherlands. 133 - 144.

Qi XY, Ren XY, An T, Chen HY, Huang J, Zhang L, 2016. Transcriptome analysis of diapause-associated genes of *Coccinella septempunctata* L. *J. Environ. Entomol.*, 38(2): 238 - 248.

Qi X, Zhang L, Han Y, Ren X, Huang J, Chen H, 2015. *De novo* transcriptome sequencing and analysis of *Coccinella septempunctata* L. in non-diapause, diapause and diapause-terminated states to identify diapause-associated genes. *BMC Genom.*, 16(1): 1086.

Robich RM, Denlinger DL, 2005. Diapause in the mosquito *Culex pipiens* evokes a metabolic switch from blood feeding to sugar gluttony. *Proc. Natl. Acad. Sci. USA*, 102(44): 15912 - 15917.

Yocum GD, Rinehart JP, Larson ML, 2009. Down regulation of gene expression between the diapause initiation and maintenance phases of the Colorado potato beetle, *Leptinotarsa decemlineata*. *Eur. J. Entomol.*, 106(4): 471 - 476.

Zhang YJ, Hao YJ, Si F, Ren S, Hu G, Shen L, Chen B, 2014. The *de novo* transcriptome and its analysis in the worldwide vegetable pest, *Delia antiqua* (Diptera: Anthomyiidae). *G3*, 4(5): 851 - 859.

Zhao JY, Zhao XT, Sun JT, Zou LF, Yang SX, Han X, Zhu WC, Yin Q, Hong XY, 2017. Transcriptome and proteome analyses reveal complex mechanisms of reproductive diapause in the two-spotted spider mite, *Tetranychus urticae*. *Insect Mol. Biol.*, 26(2): 215 - 232.

基于转录组和基因表达谱的葱蝇分泌组挖掘及其在滞育发育中的可能作用分析

郝友进*，徐燕玲，陈　斌

（重庆师范大学昆虫与分子生物学研究所，媒介昆虫重庆市重点实验室，重庆 401331）

摘要：【目的】昆虫滞育是一种重要的季节性适应生存的环境生理现象，但是调控葱蝇 *Delia antiqua* 夏滞育和冬滞育的分子机理却知之甚少。本研究旨在鉴定与葱蝇蛹滞育相关的候选基因。【方法】利用 RNA-seq 和数字基因表达谱数据，通过生物信息学的方法预测分析葱蝇滞育相关的分泌蛋白，并进行了氨基酸序列相似性、基本理化性质分析；从表达谱中选取 12 个差异表达基因，利用 qRT-PCR 验证了它们在非滞育蛹以及夏滞育和冬滞育蛹的不同发育时期的表达特点。【结果】共鉴定出 38 个在葱蝇夏滞育蛹和冬滞育蛹之间差异表达的可能的分泌蛋白。同源基因注释表明，这些基因可能在气味分子结合、几丁质和脂类代谢、免疫及发育等方面发挥作用，其中脂类和几丁质代谢过程的调节似乎在滞育发育过程起到重要作用。qRT-PCR 分析结果表明，12 个被选择基因的表达谱与 RNA-seq 结果相吻合（$R = 0.862$, $P = 0.001$）。【结论】本研究为基于高通量测序的分泌组挖掘及滞育相关基因的鉴定提供了强有力地分析策略。

关键词：葱蝇；滞育；分泌蛋白；转录组；数字基因表达谱；RNA-seq

中图分类号：Q966　　**文献标识码：**A　　**文章编号：**0454-6296(2017)10-1168-10

（责任编辑：马丽萍）